Periodicals of Engineering and Natural Sciences ISSN 2303-4521
Vol. 8, No.1, March 2020, pp.221-230

Using some classical and neural networks methods for demography predicting

Muthanna Subhi Sulaiman[1], Osamah Basheer Shukur[2]
[1]Department of Statistics and Informatics, College of Computer Science and Mathematics University of Mosul
[2] Head of Department of Statistics and Informatics, College of Computer Science and Mathematics, University of Mosul

ABSTRACT

Studying the size of populations through demography predicting may analyze the relationships between these populations and the effects of economic, biological, and social processes. In this paper, the size and the growth of world population will be predicting. Poisson and logistic models will be used as classical statistical methods, while the intelligent method will be the recurrent neural networks (RNN) technique. The results reflect that the neural networks approach outperforms the classical methods in predicting the population size. As conclusion, the neural networks approach can give better predicting accuracy than the classical statistical methods.

Keywords: Poisson model, logistic model, neural networks, demography predicting.

Corresponding Author:

Muthanna Subhi Sulaiman
Department of Statistics and Informatics
College of Computer Science and Mathematics
University of Mosul, Mosul, Iraq
E-mail: muthanna.sulaiman@uomosul.edu.iq

1. Introduction

Since 1960, the world has witnessed rapid population growth. Scientists predict that this population increase will have a clear impact on social and economic life in the coming years, as the demand for food and water in particular will increase this population growth is called population growth. Known as the increase in population during a certain period, and called the word population growth refers to the change in the size of the population, whether this change in the increase or decrease, which represents births, deaths and migration.
Population modeling and predicting are the main problems that ecology researchers are interested in. Population modeling involves studying changes in population sizes as a result of interactions between individuals in the natural environment with individuals of their own kinds. Population modeling also aims to study the growth or decay in the numbers of humans and animal organisms in the communities in which they live. There are many common mathematical tools in this regard. The estimation of the population is an important issue that different world areas are interested in. Often a census is conducted every ten years to find out the real population numbers [1].
This paper aims to study the issue of modeling the global population size to study population growth and forecast the population numbers in the future. Poisson and logistic models will be used as classical statistical methods, while the intelligent method will be the neural networks approach.

2. Methods and materials

Population modeling aims to study the growth or decay of populations in the communities in which they live. There are many traditional statistical tools based mainly on theoretical assumptions such as Poisson model and logistic model. The neural network will be used also for predicting as one of artificial intelligent approaches.

2.1. Exponential model

The first researcher who suggests using the exponential model in the population is Thomas Robert Malthus. The basic assumption of this model is that the numbers of births and deaths are proportional to both the size of the community and the length of time during short period.

Assuming that $N(t)$ represents the size of population of a given region in time t, $B(t)$ represents the number of births in that region in time t, and $D(t)$ represents the number of deaths in that region in time t. If δ_t represents a small period of time, the mathematical model of the previous hypothesis would be as follows [2].

$$B(t) = \alpha N(t)\delta t \tag{1}$$

$$D(t) = \beta N(t)\delta t \tag{2}$$

where, α and β represent constant values.

The change of population number in a small period of time represents the difference between births and deaths in that period according to the basic assumption of this model.

$$N(t) = N(0)e^{ct} \tag{3}$$

$N(t)$ represents the size of the population in year t, and $N(0)$ is the size of the population in the base year. And c represents the growth rate in the population and is calculated from the following relationship [3,4]:

$$c = \ln\left[\frac{N(t)}{N(0)}\right] \tag{4}$$

2.2. Logistic model

This model was proposed by researcher Pierre François Verhulst. The overall form of the logistic model is as a sigmoid curve that describes population growth exponentially. Assume the carrying capacity $N(\infty)$ to be the maximum capacity of the population $N(t)$ representing the size of the population in time t. therefore, the logistic model will be as follows.

$$\frac{dN(t)}{dt} = cN(t)\left(1 - \frac{N(t)}{N(\infty)}\right) \tag{5}$$

$$N(t) = \frac{N(\infty)}{\left(1 + \left(N(\infty)/N(0) - 1\right)e^{-ct}\right)} \tag{6}$$

$$c = -\ln\left(\frac{N(\infty) - N(1)}{N(1)\left(N(\infty)/N(0) - 1\right)}\right) \tag{7}$$

Root mean square error (RMSE) measurement is computed for the error for all methods and all datasets as a statistical criterion to evaluate the adequacy and accuracy of these methods [5, 6].

2.3. Recurrent neural network (RNN) approach

Hu [7] was the first that investigated the weather forecasting using an artificial neural network (ANN). A lot of studies came later to investigate using ANN for forecast the growth of population. In this study, the recurrent algorithm of neural network to enhance the forecasting accuracy because this algorithm has a delay layer that gives longer memory comparing to other algorithms. Longer memory is an advantage that handles the heteroscedasticity and the trend of population growth. The general structure of RNN can be described such as follows.

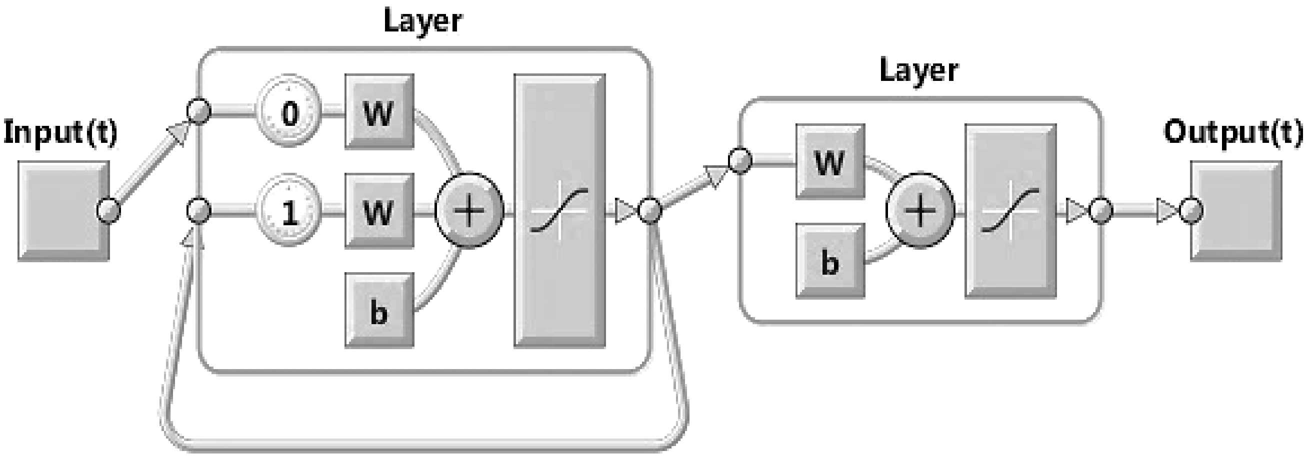

Figure 1. The general structure of RNN

In this study, RNN will be such as in figure above. It includes 1 input layer, 1 hidden layer, and 1 output layer. There are several input variables in the input layer and several neurons in the hidden layer, while it is preferable for output layer to be with one output variable only. Let the number of input variables be (R), then the preferable number of neurons is ((R×2)+1) [8, 9]. Each input variable is separately weighted by a random weight w automatically. The weighted inputs for R variables and M neurons are summed with the biased value b to formulate the input of the transfer function. The input variable SUM of the transfer function f can be formulated such as follows.

$$SUM = \sum_{i=1}^{M} \sum_{j=1}^{R} w_{i,j} Z_j + b. \tag{8}$$

The most used types of transfer functions for hidden and output layers are commonly tan-sigmoid that generates outputs between -1 and +1, log-sigmoid that generates outputs between 0 and 1, and linear transfer functions that generates outputs between -1 and +1. Selecting a transfer function for the hidden and output layers is important for obtaining good results and depends on the nature of the data. Figure 2 displays the different types of transfer function that can be specified using ANN toolbox in MATLAB.

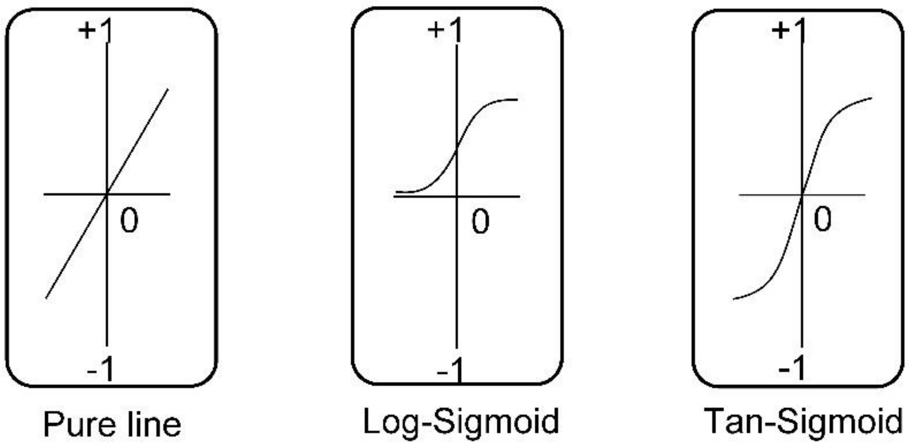

Figure 2. RNN transfer function types.

The transfer function in the hidden layer is used to reflect the type of the relationship between the input and output variables, while the transfer function in the output layer is commonly employed to constrain the observation values of RNN's output variable. The mathematical expressions for linear, logistic sigmoid, and tangent sigmoid transfer functions, respectively, are as follows.

$$f(SUM) = SUM \tag{9}$$

$$f(SUM) = \frac{1}{1+e^{-SUM}} \tag{10}$$

and $$f(SUM) = \frac{2}{1+e^{-2SUM}} - 1 \tag{11}$$

Where SUM was defined in equation (9) [10-12]. The random weights $w_{i,j}$ for R inputs and M neurons can be written as a matrix such as follows.

$$W = \begin{bmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,R} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,R} \\ \vdots & \vdots & \ddots & \vdots \\ w_{M,1} & w_{M,2} & \cdots & w_{M,R} \end{bmatrix}, \tag{12}$$

While the inputs vector can be formulated as follows.

$$Z = \begin{bmatrix} Z_{t1} & Z_{t2} & \dots & Z_{tR} \end{bmatrix}' \tag{13}$$

RNN training process is employed to adjust the random weights and bias values in each layer to enhance the forecasting accuracy and obtain desirable output. Training process may be supervised or unsupervised.

3. Results and discussions

In this study, modeling and prediction of the growth of global population size will be performed using MATLAB software. The application includes using Poisson and logistic models as classical statistical methods, and also includes using RNN as intelligent approach.

RNN approach has been performed using the neural network toolbox in MATLAB after preparing input and target variables for training and testing separately. The structure of input variables are suggested to be the Z_t's lagged series that represent the autoregressive variables such as used by [13-18]. There is one target variable and on corresponding output variable that can be compared to obtain a prediction error and mean absolute percentage error (MAPE).

Many accuracy measurements were employed in previous studies to quantify the error of prediction such as root mean square error (RMSE), MAPE, and others. RMSE is commonly used as the dispersion measurements, while MAPE is commonly used to scale the percentage prediction error and the accuracy of prediction [19, 20]. Therefore, only MAPE can be used in this study sufficiently. MAPE is written such as follows.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{e_i}{Z_i}\right| \times 100 \tag{14}$$

Where e_i is the forecasting error, *n* is the number of observations, and Z_i is the original series that is used as a target.

The growth of global population size dataset is taken for the period (1961 till 2002). This period is divided into two groups, the first one is for training for the period (1961 till 1995) while the second is for testing for the period (1996 till 2002). Full datasets is such as in following table.

Table 1. The global population size full data for (1961 till 2002)

t	Years	Population size	*t*	Years	Population size	*t*	Years	Population size
1	1961	3080	15	1975	4086	29	1989	5194
2	1962	3136	16	1976	4158	30	1990	5282
3	1963	3206	17	1977	4230	31	1991	5365
4	1964	3277	18	1978	4302	32	1992	5449
5	1965	3346	19	1979	4377	33	1993	5531
6	1966	3416	20	1980	4453	34	1994	5611
7	1967	3486	21	1981	4529	35	1995	5692
8	1968	3558	22	1982	4608	36	1996	5771
9	1969	3632	23	1983	4690	37	1997	5850
10	1970	3707	24	1984	4770	38	1998	5928
11	1971	3785	25	1985	4852	39	1999	6004
12	1972	3862	26	1986	4935	40	2000	6080
13	1973	3938	27	1987	5021	41	2001	6154
14	1974	4013	28	1988	5107	42	2002	6227

Source: https://www.census.gov/popclock

The figures of the training and testing datasets are as follows:

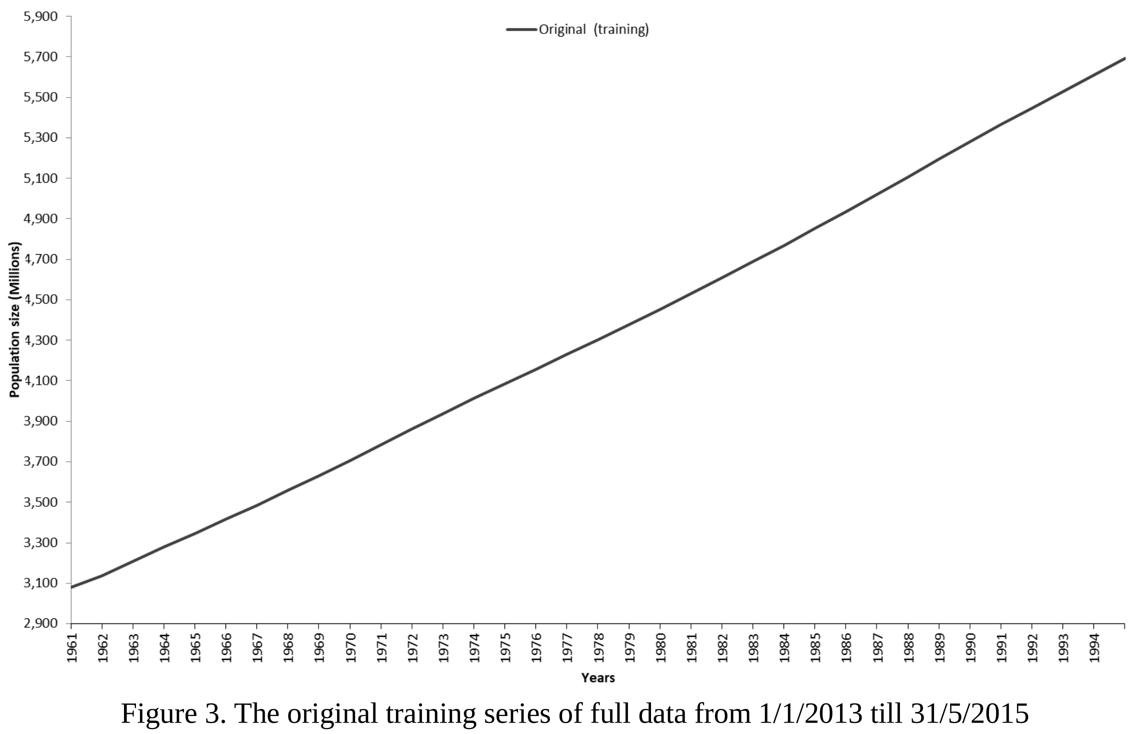

Figure 3. The original training series of full data from 1/1/2013 till 31/5/2015

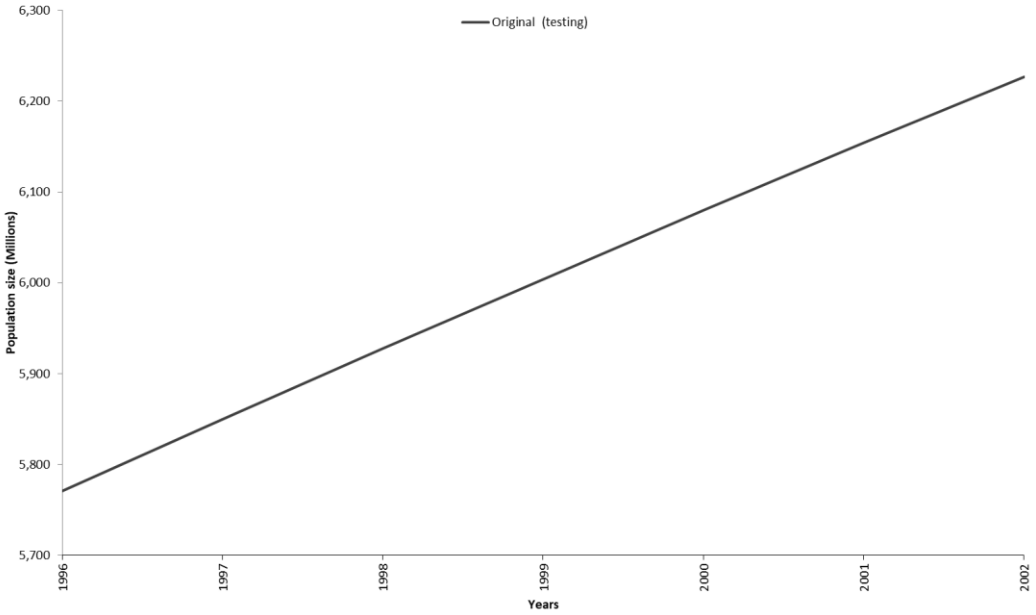

Figure 4. The original testing series of full data from 1/6/2013 till 31/10/2015

The details of the training and testing forecasts for global population sizes using MAPE to reflect the accuracy of prediction results are displayed such as in following table.

Table 2. MAPE of the training and testing forecasts for global population sizes datasets

Method	**MAPE**	
	Training	Testing
Exponential	1.85	1.53
Logistic	8.74	7.42
RNN	0.26	0.03

The fitness figures of the training and testing datasets between original and forecast variables for all used method are such as follow.

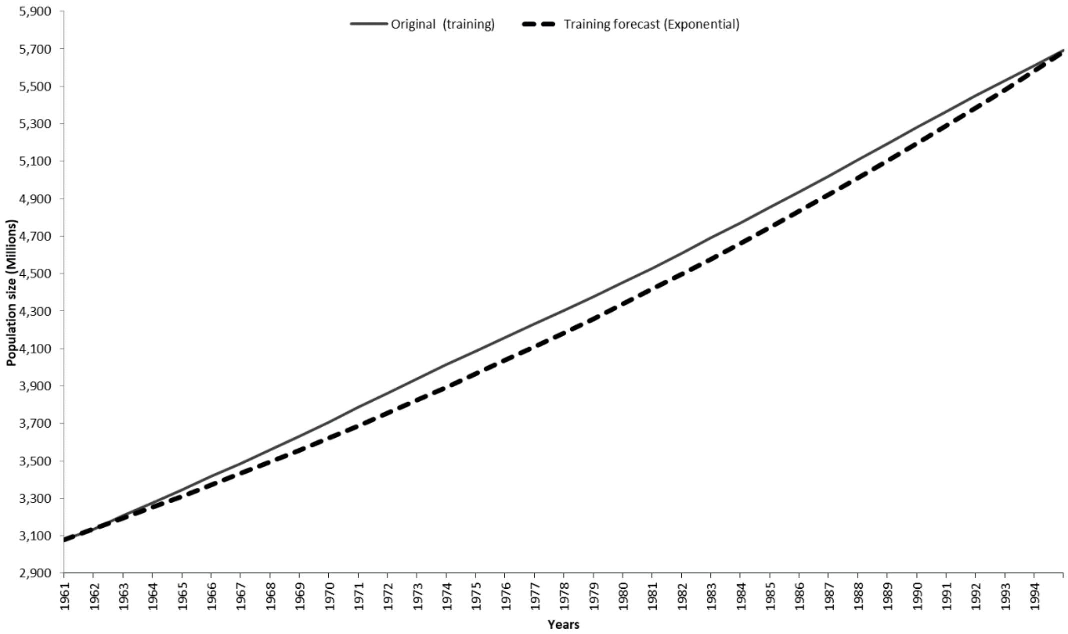

Figure 5. The fitness of original training series using exponential method.

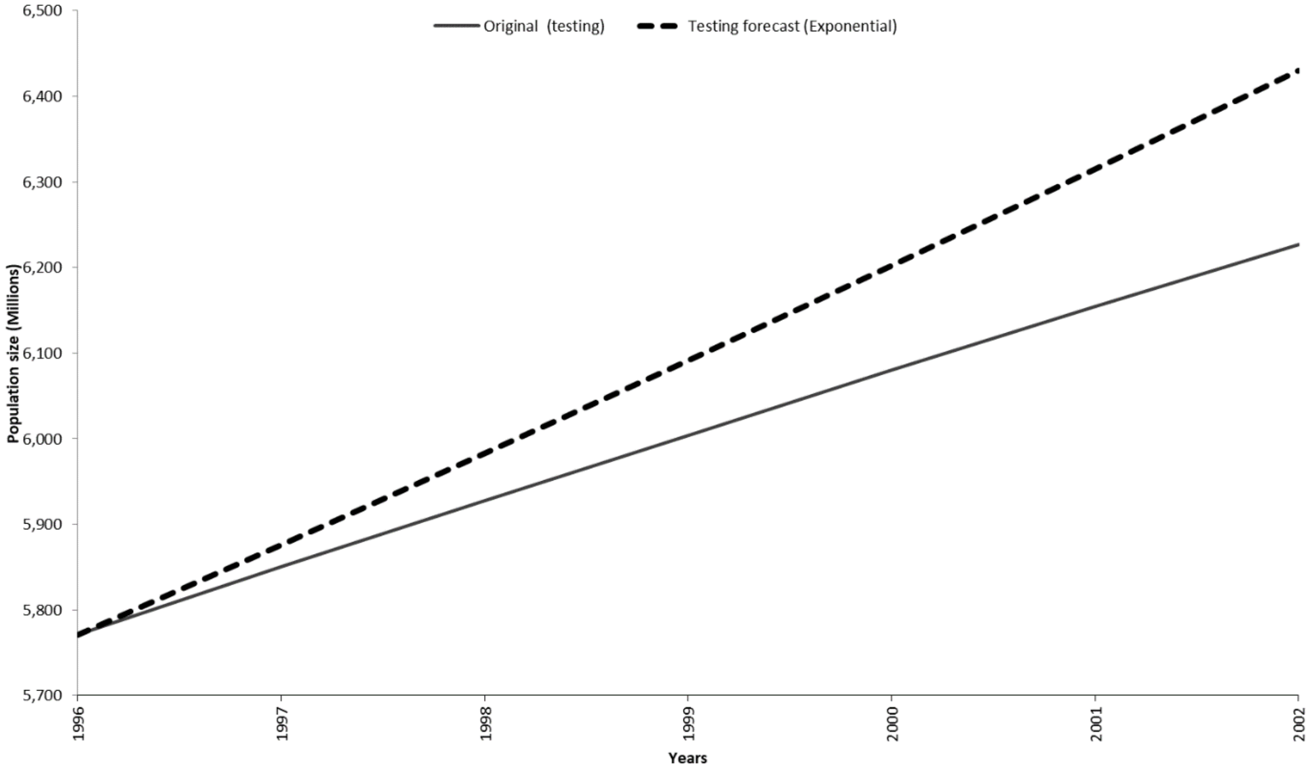

Figure 6. The fitness of original testing series using exponential method

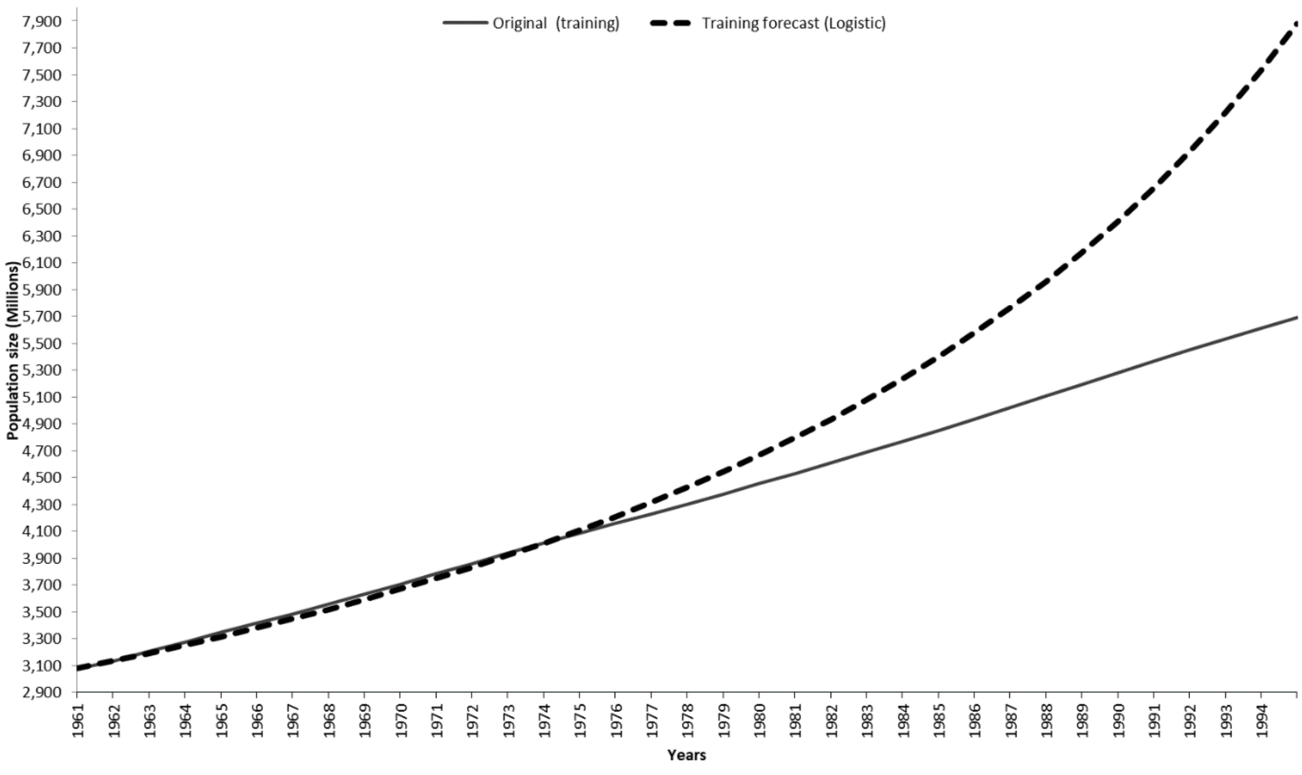

Figure 7. The fitness of original training series using logistic method.

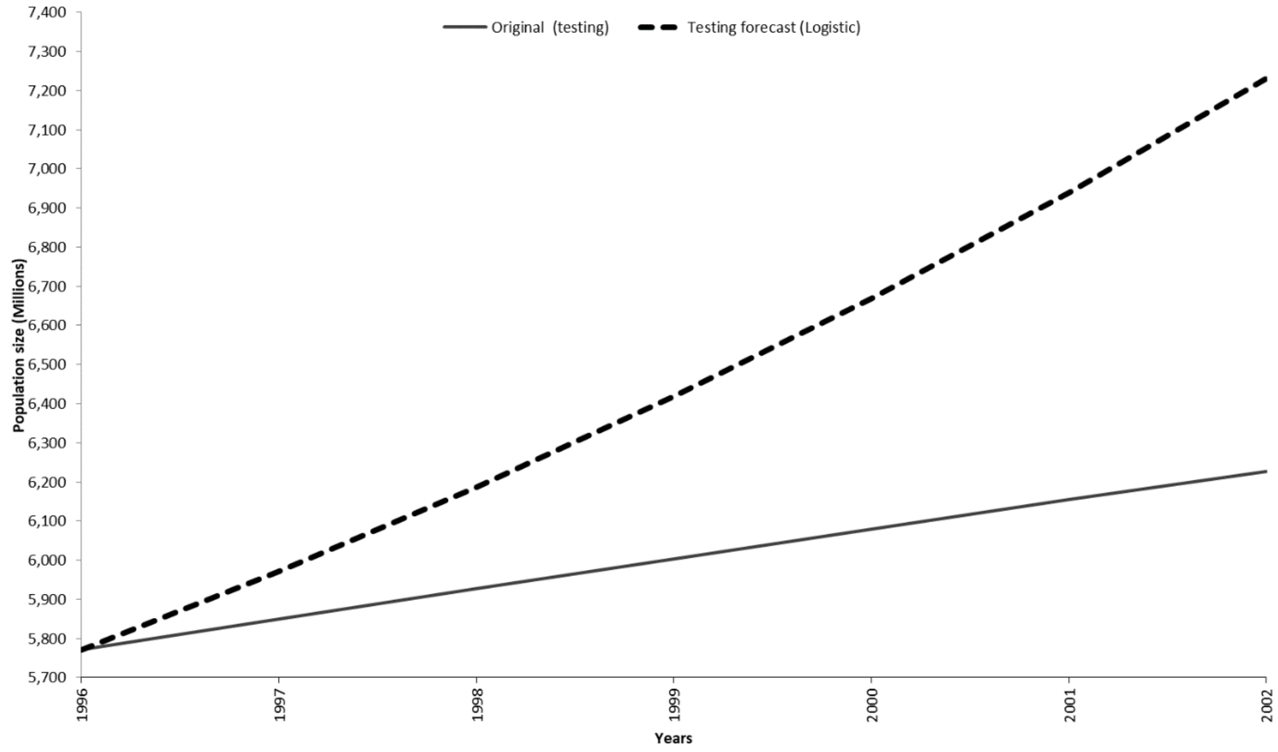

Figure 8. The fitness of original testing series using logistic method

Figure 9. The fitness of original training series using RNN method.

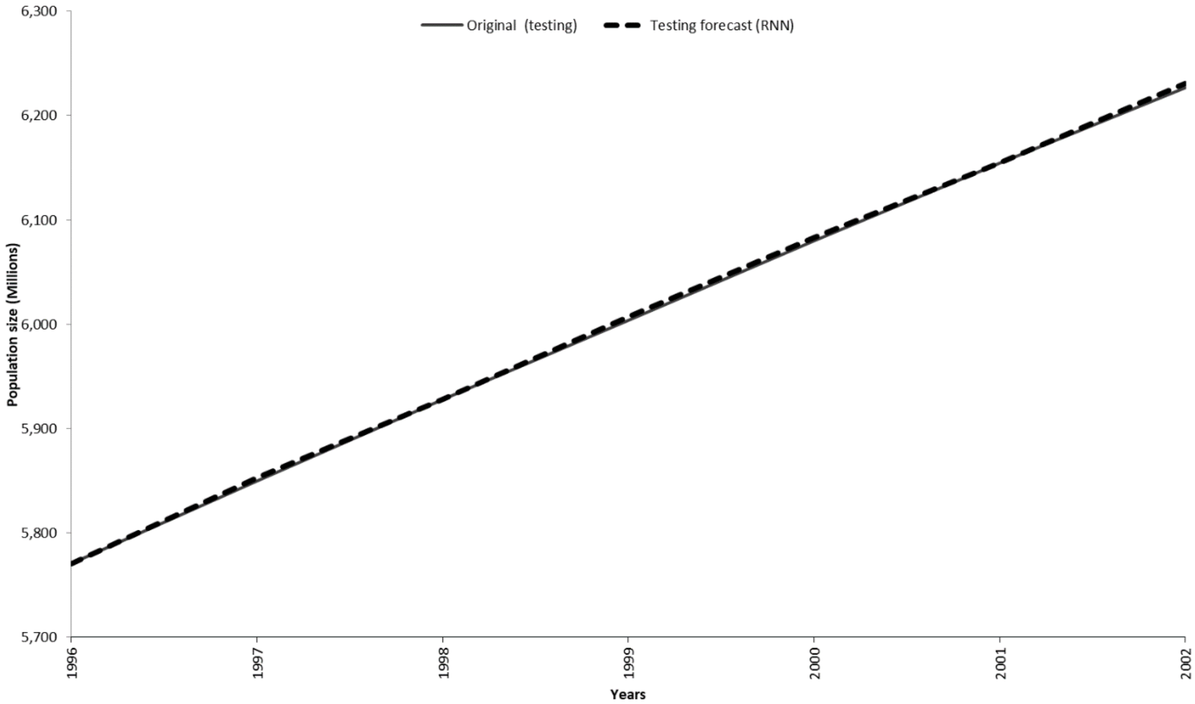

Figure 10. The fitness of original testing series using RNN method

From all figures and tables of the training and testing forecasts for global population sizes, the accuracy prediction and fitness between original and forecast series using RNN outperformed the results of classical methods.

4. Conclusions

After performing the training and testing forecasts for global population sizes, RNN can be used for predicting the growth of global population sizes in high accuracy prediction and it results will outperform the results of other classical statistical methods.

5. Acknowledgments

The authors are very grateful to the University of Mosul/ College of Computer Sciences and Mathematics for their provided facilities, which helped to improve the quality of this work.

References

[1] L.R. Ginzburg, et al., "Quasiextinction probabilities as a measure of impact on population growth" *Risk Analysis*,.vol. 2, no.3, pp. 171-181, 1982.

[2] A. Costamagna, et al., *"An exponential growth model with decreasing r captures bottom- up effects on the population growth of Aphis glycines Matsumura (Hemiptera: Aphididae)", Agricultural and Forest Entomology*, vol. 9, no.4, pp. 297-305, 2007.

[3] J. Mohanty, B. Verma, and P. Ray, Prediction of fatigue crack growth and residual life using an exponential model: Part I (constant amplitude loading), I*nternational Journal of Fatigue*, vol.31, no.3, pp. 418-424, 2009.

[4] J. Mohanty, B. Verma, and P. Ray, Prediction of fatigue crack growth and residual life using an exponential model: Part II (mode-I overload induced retardation), *International Journal of Fatigue*, vol. 31, no.3, pp. 425-432, 2009.

[5] S. Sakanoue, Extended logistic model for growth of single-species populations, *Ecological modelling*, vol. 205, no.(1-2), pp. 159-168, 2007.

[6] M. Ferrara and L. Guerrini, The Ramsey model with logistic population growth and Benthamite felicity function revisited, *WSEAS Transactions on Mathematics*, vol. 8, no.3, pp. 97-106, 2009.
[7] M.J.-C. Hu, *Application of the adaline system to weather forecasting*, Department of Electrical Engineering, Stanford University. Stanford, CA, 1964.
[8] Palit, A.K. and D. Popovic, *Computational intelligence in time series forecasting: theory and engineering applications*, Springer, 2006.
[9] K.G. Sheela and S. Deepa, Review on methods to fix number of hidden neurons in neural networks, *Mathematical Problems in Engineering*, vol.2013, 2013.
[10] H. Yonaba, F. Anctil, and V. Fortin, Comparing sigmoid transfer functions for neural network multistep ahead streamflow forecasting, *Journal of Hydrologic Engineering*, vol.15, no.4, pp. 275-283, 2010.
[11] C. Dawson and R. Wilby, Hydrological modelling using artificial neural networks, *Progress in physical Geography*, vol. 25, no. 1, pp. 80-108, 2001.
[12] R.R. Shrestha, S. Theobald, and F. Nestmann, Simulation of flood flow in a river system using artificial neural networks, *Hydrology and Earth System Sciences*, vol. 9, no.4, pp. 313-321, 2005.
[13] H. Liu, H.-q. Tian, and Y.-f. Li, Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction, *Applied Energy*, vol. 98, pp. 415–424, 2012.
[14] O.B. Shukur and M.H. Lee, Daily Wind Speed Forecasting Through Hybrid AR-ANN and AR-KF Models. *Jurnal Teknologi*, vol. 72, no.5, pp. 89–95, 2015.
[15] G.P. Zhang, Time series forecasting using a hybrid ARIMA and neural network model, *Neurocomputing*, vol. 50, pp. 159-175, 2003.
[16] M. Khashei and M. Bijari, An artificial neural network (p,d,q) model for time series forecasting, *Expert Systems with Applications*, vol. 37, no.1, pp. 479-489, 2010.
[17] M. Khashei and M. Bijari, A novel hybridization of artificial neural networks and ARIMA models for time series forecasting, *Applied Soft Computing*, vol. 11, pp. 2664–2675, 2011.
[18] O.B. Shukur, et al., Electricity Load Forecasting using Hybrid of Multiplicative Double Seasonal Exponential Smoothing Model with Artificial Neural Network, *Jurnal Teknologi*, vol. 69, no.2, pp. 65-70, 2014.
[19] B.L. Bowerman, R.T. O'Connell, and A.B. Koehler, *Forecasting, time series, and regression : an applied approach*, Belmont, Calif: Thomson Brooks/Cole Cengage Learning, 2005.
[20] R.J. Hyndman and A.B. Koehler, Another look at measures of forecast accuracy, *International Journal of Forecasting*, vol. 22, no.4, pp. 679-688, 2006.
[21] H. Peng, F. Liu, and X. Yang, A hybrid strategy of short term wind power prediction, *Renewable Energy*, vol. 50, pp. 590-595, 2013.